DOCUMENT RESUME

ED 335 015 IR 015 182

AUTHOR Stepp, Sidney Leland; Shrock, Sharon A.
TITLE The Validity of a Multiple-Choice, Paper and Pencil Instrument in Discriminating between Masters and Nonmasters of Instructional Design.
PUB DATE 91
NOTE 27p.; In: Proceedings of Selected Research Presentations at the Annual Convention of the Association for Educational Communications and Technology; see IR 015 132.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Competence; Discriminant Analysis; Higher Education; *Instructional Design; Predictor Variables; *Test Validity
IDENTIFIERS *Paper and Pencil Tests

ABSTRACT

While surveys have identified instructional design competencies, there has been virtually no systematic research of alternative means for assessing professional competence in this area. This paper reports on a study which investigated the question of whether a multiple-choice, paper and pencil test can validly discriminate between levels of professional competency in instructional design. The instructional design instrument was developed in three stages: (1) items were composed and revised until subject matter experts agreed to each item's logical validity; (2) trial testing and item analysis were done using groups of masters and nonmasters of instructional design to test empirically and eliminate non-discriminating items from the instrument; and (3) a phi coefficient was calculated to show a level of concurrent validity, the Tukey method of multiple range means testing was used to show significant differences between groups and subgroups, and multiple discriminant function analysis was used to identify additional items that did not discriminate between groups. A total of 257 subjects completed the instrument. A comparison of the mastery classification of the subject matter experts and the classification of instrument established the concurrent validity of the instrument, and it was concluded that this type of instrument can validly discriminate between masters and nonmasters of instructional design although further research is needed. A list of references used to identify the competencies and sample questions from the instrument are appended. (5 tables, 8 figures, and 37 references) (BBM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED335015

Title:

The Validity of a Multiple-Choice, Paper and Pencil Instrument in Discriminating between Masters and Nonmasters of Instructional Design

Authors:

Sidney Leland Stepp
Sharon A. Shrock

IR015182

BEST COPY AVAILABLE

The Validity of a Multiple-Choice, Paper and Pencil Instrument in Discriminating Between Masters and Nonmasters of Instructional Design

Instructional design is a big business in industrial organizations. As the instructional design profession continues to grow and greater investments of time and money are made, instructional designers may be held more and more accountable for the instructional decisions made and programs developed. Questions of ability or selection will become more prominent and instruments that can validly be used in assisting to make judgements will become a paramount issue.

Assessing competency in instructional design has been a hot topic for debate for a number of years. While surveys have identified instructional design competencies, there has been virtually no systematic research of alternative means for assessing professional competence. The purpose of this paper is to report a study investigating the question: Can a multiple-choice, paper and pencil test validly discriminate between masters and nonmasters of instructional design?

The Issues Involved

In considering the study two sets of issues emerge: psychometric and political. The psychometric issues involve test item formats and the nature of the instrument. Seven item formats were considered: true/false, matching, fill-in-the-blank, short answer, essay, multiple-choice, and performance in an assessment center. Of the seven, actual performance in an assessment center is the most valid form of identifying instructional design competencies although assessment centers would be expensive in terms of the time required to perform the assessments and money required to establish and implement the appropriate testing. These problems make the assessment center approach impractical in most circumstances. Other item formats are inappropriate because of the inability of the formats to get at higher cognitive levels or difficulties in achieving scoring reliability. A multiple-choice instrument would overcome the problem of expense, provided that a valid, discriminating instrument could be developed.

The choice between a norm- and criterion-referenced instrument is difficult. Politically, the field of instructional design would most readily accept a criterion-referenced instrument or at least a norm-referenced instrument that covers all of the competencies found important for an instructional designer. Unfortunately, this is not possible with this type of instrument. The instrument cannot cover all of the competencies such as consulting skills or writing ability. A statistically validated, norm-referenced instrument was selected as the best choice for a multiple-choice instrument.

Political issues involve the different focuses between instructional design organizations and professionals. For more than the past 20 years, the issue of instructional design testing and certification has been problematic at both the organizational and individual levels. While organizations express an interest in the certification issue and competency testing (Prigge, 1974), no progress and few efforts have been made as demonstrated in the literature. Organizations have not united to work together perhaps due to their varying special interests and audiences. In the single case where people actually sat down together to develop a criterion-referenced test with objectives-based test items (NSPI/DID-AECT Task Force in 1982), the varying backgrounds of the group provided an overwhelming stumbling block (Sharon A. Shrock, personal communication, May 10, 1989). Failed attempts aside, the research has not

addressed the question: Can valid items be constructed that can make fine discriminations for an instructional design competency instrument?

The Three Stages of Instrument Development

The instructional design instrument was developed in three stages illustrated in Figure 1. First, items were composed and revised until subject matter experts agreed to each item's logical validity. Second, trial testing and item analysis were done to test empirically and eliminate non-discriminating items from the instrument. In this stage two groups of subjects contributed to the data: one group of non-professional instructional design masters and one group of instructional design nonmasters. In the third stage a phi coefficient was calculated to show a level of concurrent validity. The Tukey method of multiple range means testing was used to show significant differences between groups and subgroups. Multiple discriminant function analysis was used to identify additional items that did not serve to discriminate between groups. In this third stage two new groups contributed to the data: professional masters and nonmasters. In addition, the nonmasters group was divided into four subgroups: education graduate students, education undergraduate students, non-education graduate students and non-education undergraduate students. The nonmasters were split into these four subgroups to obtain information concerning the instrument's ability to make fine discriminations between overlapping groups.

Composing and Revising Items

Item composition began following a brainstorming session with subject matter experts. At that meeting the specific means of item writing and stages of item analysis were discussed. Several conclusions were reached.

Resources for the Item Bank

The multiple-choice items for the item bank were written. Non-knowledge level principles and concepts (as classified by Bloom's and Gagne's taxonomy) used for the items were selected from instructional design textbooks listed in Appendix A.

The 50 item bank and subject matter expert review. The items were reviewed by subject matter experts. During the meetings, the subject matter experts discussed item clarity, ambiguity, and logical validity. Suggestions for item changes or removal were made. A total of 35 items were agreed upon for the instrument's item analysis.

Trial Testing and Item Analysis

Groups of masters and nonmasters of instructional design were identified by the subject matter experts. After volunteering to participate, the master and nonmaster groups were asked to complete a demographic data sheet and the instrument.

Non-Professional Masters

A total of 17 current and former graduate students in the Department of Curriculum and Instruction at Southern Illinois University were identified by the subject matter experts as having a mastery of instructional design. The mastery decision was based on the subject matter experts' observations of the students' course work in instructional design, work with clients, and interactions in the classroom. Of this group, 16 students completed the instrument.

Nonmasters

A total of 57 nonmasters of instructional design were identified by the subject matter experts. The selection of the nonmasters from these particular classes was based

on the subject matter experts' knowledge of the students and their abilities (or inabilities) in instructional design rather than the courses taken by the students in instructional design.

Item Analysis

Group means for the 73 non-professional masters and nonmasters are illustrated in Figure 2. Pearson point-biserial coefficients ranged from -.14 to .41 for the instrument items. Seven items were removed from the instrument that had negative Pearson point-biserial coefficients. Item analysis was performed again using the data from the 28 remaining items. Pearson point-biserial coefficients ranged from .09 to .54 with no negative Pearson point-biserial coefficients remaining in the data set. The Cronbach alpha coefficient increased from .5521 for the 35 item instrument to .6574 for the 28 item instrument.

Three other item analyses were performed removing other items with low Pearson point-biserial coefficients to find if results could be further improved. No other analysis provided better results than the initial 28 item analysis. The 28 item instrument was selected for further data collection and instrument validation.

Concurrent Validity

At this point in the research new data was collected with the 28 item instrument for instrument validation. To enhance the generalizability of the data analysis results it was important to broaden the population used to include subjects outside the College of Education at Southern Illinois University and outside the Southern Illinois University environment. New groups of masters and nonmasters of instructional design were identified and solicited to complete the instrument.

A total of 48 professional masters of instructional design concepts were identified from the membership of the National Society for Performance and Instruction and the Association for Educational Communications and Technology. Packages containing the instrument materials, an addressed, stamped envelope, and a letter briefly describing the research were sent to the 48 professional masters.

Professional Masters

The 48 professional masters (PM) were selected for their known expertise in the field of instructional design. Geographically, the group was spread across the United States and Canada. A total of 34 completed instruments and three incomplete instruments were returned providing a total response rate of more than 77%.

Demographic data. While 34 completed instruments were returned, demographic data sheets were not returned for two of the instruments. The following data reflect the 32 demographic data sheets that were returned. Also, some subjects did not complete all items on the demographic data sheets. The following data reflect completed items for the sheets that were returned.

Table 1 shows a summary of the demographic data concerning the respondents' jobs collected from the professional masters. As can be seen from the table, respondents stated that they held one of six types of positions ranging from academic faculty to business executive to private consultant. Respondents were equally split between academic and business affiliation with 16 respondents from each of the two types of positions. Table 2 shows the degrees earned by the professional master respondents. The majority (29) responded that they had a Ph.D. or an Ed.D. while 1 held an M.A. and 1 held a B.A. All but one subject responded that jobs held were instructional design related. A total of 28 respondents indicated that positions held were related to

education while 3 indicated that positions held were not educationally oriented. In all three of these cases the job held was corporate management.

Table 3 shows the programs where the degrees were obtained by the professional masters. Eleven different programs were indicated. The largest number of respondents indicated that they obtained their degrees in programs called instructional systems technology. Additionally, not all programs were directly related to instructional design such as anthropology and psychology.

Finally, the range of hours of course work in instructional design taken by the professional masters varied greatly. A total of 19 respondents indicated that they had taken more than 12 hours, 3 respondents indicated 6 to 12 hours, 1 respondent indicated 3-6 hours, and 8 respondents indicated less than 3 hours. The relatively large number of subjects responding that they had taken less than 3 hours could be due to the newness of instructional design programs combined with the ages of the respondents (i.e., there are three types of people in the field of instructional design: experience without instructional design education, instructional design education without on-the-job experience, and experience with instructional design education.) One respondent wrote a note indicating that the respondent's degree was obtained before courses in instructional design were offered in the respondent's program. Respondents were asked to complete the instrument because of their known expertise in the field. In some cases this expertise was based on years of work experience. Some respondents had not attended recently developed programs of instructional design to learn theories about what they were already practicing and in many cases publishing in the field.

Nonmasters

The nonmasters group was composed of students enrolled at Southern Illinois University. The group contained four subgroups: 43 education graduate students (EGS), 45 education undergraduate students (EUS), 23 non-education graduate students (NEGS), and 39 non-education undergraduate students (NEUS).

Descriptive Analysis

Table 4 shows the descriptive analysis of the instrument responses for the 184 professional masters and nonmasters in the study. The table shows the professional master group and nonmaster subgroup, ranges, means, and standard deviations. As can be seen in the table, the professional master group and four nonmaster subgroup numbers range from 23 to 45, the means range from 7.949 to 19.265, and the standard deviations range from 2.516 to 3.336. Figure 3 illustrates the various mean scores for the professional master group and the four nonmaster subgroups.

Setting the Cut Off Level and Demonstrating Concurrent Validity

Figure 4 shows the smoothed frequency distributions of scores for the professional masters overlaid with the frequency distribution of scores for the nonmasters. The figure shows that the distributions for the two groups intersect at a score of 17. As described by Allen and Yen (1979) this intersection of frequencies can be accepted as the cut off level for computing phi to demonstrate the validity of the instrument. The consequences of choosing a score of 17 as the cut off level are two-fold. First, some identified masters will be misclassified by the instrument as nonmasters. Second, some identified nonmasters will be misclassified as masters by the instrument. The main objective in setting this cut off score is to minimize the misclassification for both consequences.

A phi correlation was used to demonstrate the instrument's concurrent validity. Figure 5 shows the arrangement of data used for the calculation. The figure also shows the percent of the master/nonmaster groups falling into each of the four categories. The four categories are formed from the master/nonmaster classification of the subject matter experts and the master/nonmaster classification of the instrument at a cut off level of 17. The figure shows that the instrument would misclassify 4 masters or 11.76% of the master group and 16 nonmasters or 10.67% of the nonmaster group.

At a cut off level of 17 the phi coefficient produced was .695. Subsequent runs with other cut-off levels (14, 15, 16, 18, 19, and 20) did not produce a greater phi coefficient demonstrating the validity of the choice of 17 as the cut off level. At a cut off level of 16 a phi coefficient of .601 was produced. At this level approximately 6% of the masters would be misclassified and approximately 21% of the nonmasters would be misclassified. At a cut off level of 18 a phi coefficient of .626 was produced. At this level approximately 7% of the nonmasters and about 29% of the masters would be misclassified.

Tukey Method of Multiple Range Means Testing

Table 5 shows the results from running the Tukey Method of multiple range testing of the five groups/subgroups of means. As is shown in that table, using a harmonic N ($\underline{N}$ = 34.72) to obtain an average sample size, the Tukey method would require a 1.82 difference between mean scores for significance at $\underline{p} < .05$. The table shows that the mean score of the professional masters was significantly higher than the mean score of students in education, the mean score of students in education was significantly higher than the mean score of graduate students not in education, and the mean score of graduate students not in education was significantly higher than undergraduate students not in education. What is surprising in the results is that the mean scores of graduate and undergraduate students in education were *not* significantly different. At the same time the mean score of undergraduates not in education was significantly lower than the graduate students not in education. Differences between mean scores do not appear to exist because of differences between graduate and undergraduate abilities.

A Look at the Individual Items and Multiple Discriminant Function Analysis

Multiple discriminant function analysis was performed using the 28 instrument items as group membership predictors to identify any items that did not predict group membership. The five groups used in the analysis were professional instructional designers (PM), education graduate students (EGS), education undergraduate students (EUS), non-education graduate students (NEGS), and non-education undergraduate students (NEUS). The univariate F tests demonstrated that all but four of the items significantly discriminated between groups, $\underline{p} \leq 0.05$. The Wilk's Lambda, also computed in the analysis, demonstrated similar results to the Tukey multiple range analysis. The groups do differ significantly (aside from the education graduates and education undergraduates) on all instrument items as a set. The Wilks' lambda was calculated to be .063. This is equivalent to a statistically significant $\underline{F}(112, 606) = 5.458$, $\underline{p} \leq .05$.

On the basis of all 28 predictors, Chi square tests were computed for each of the four derived discrimination functions (based on five groups minus one) to determine the significance of discrimination along each dimension. The first discriminant function was found to be significant $\underline{x}^2(112, \underline{N} = 184) = 461.09$, $p \leq .05$. The second and third functions were also found to be significant $\underline{x}^2$ (81, $\underline{N}$ = 184) = 200.58, $\underline{p} \leq .05$ and $\underline{x}^2(52, \underline{N} = 184) = 96.149$, $\underline{p} \leq .05$. The fourth dimension failed to reach the necessary level

for significance (p = .30). The first three discrimination functions accounted for 97% of the variance between groups. Greater coefficient values show a greater ability of the item to discriminate between groups. Four items on the instrument had low values. The univariate analysis indicated that these items did not help the instrument discriminate between the five groups.

Figure 6 shows the discriminant analysis results for the predicted versus actual classification for the total sample. The diagonal line in the figure underlines those numbers of correct classifications. The high number of correct classifications demonstrates the ability of the whole instrument to discriminate between groups.

New Phi Correlation After Removing Four Items

The four items that discriminant analysis identified as not useful in discriminating between groups were removed in order to recalculate a new phi coefficient using only the items that did statistically work to discriminate between groups. New professional master group and nonmaster subgroup mean scores are shown in Figure 7. All group and subgroup mean scores are lower than the means shown in Figure 3 although it must be remembered that the mean scores shown in Figure 7 are based on a 24 item instrument while those in Figure 3 are based on a 28 item instrument. The phi coefficient after removing the four items increased from .695 to .758. The Cronbach alpha coefficient also increased from .746 to .762. Removing the four items statistically increased the validity and reliability of the instrument.

Conclusions

A total of 257 subjects completed the instrument. Item analysis and instrument validation were performed on the data. Several conclusions can now be drawn concerning the instrument, its validation, and its future.

The Validity of the Instrument

A test of concurrent validity compares mastery classifications. A comparison between the mastery classification of the subject matter experts and the classification of the instrument at a cut off level of 17 produced a phi coefficient of .695. The instrument's concurrent validity in this study has been established.

Univariate F tests in a discriminant analysis. Univariate F tests during discriminant analysis of the instrument items showed that 4 items did not serve to significantly discriminate between groups. The removal of the 4 items increased the phi coefficient produced by the data from .695 to .758. The Cronbach alpha coefficient was also increased from .746 to .762 after the removal of the 4 items. The instrument can be further refined in future studies.

Wilks' lambda in a discriminant analysis. Before discriminant functions could be generated, the five groups of data needed to be tested to see if they differed significantly on the 28 instrument items as measured by the Wilk's lambda statistic. The Wilks' lambda was calculated to be .063. This is equivalent to an F (112,606) of 5.458, $p \leq .05$. The instrument items do discriminate among the five groups.

Discriminant analysis. Chi square tests were computed for the derived discrimination functions to determine the significance of discrimination along each of the four dimensions. The first three discriminant functions were found to be significant ($\chi^2(112, N = 184) = 461.09$, $p \leq .05$; $\chi^2(81, N = 184) = 200.58$, $p \leq .05$; and $\chi^2(52, N = 184) = 96.149$, $p \leq .05$), but the significance of the fourth dimension failed to reach the necessary level (p = .30).

A Comparison of Professional and Nonprofessional Masters' Mean Scores

The mean scores of the non-professional masters used in the item analysis stage and the professional masters used in the validation stage are illustrated in Figure 8. The mean score of the professional masters ($\underline{M}$ = 19.265) was not significantly different from the nonprofessional masters ($\underline{M}$ = 19.5). In view of the years of real world experience of the professional masters and the inexperience of the nonprofessional masters this non-significant result might seem strange. Experience would seem to add to the ability of the professional masters over a simple knowledge of the theories studied by the nonprofessionals without the real world experiences. Since the profession of instructional design is relatively new and studies comparing the knowledge bases of experienced professionals and non-experienced professionals do not exist, an analogy to the medical profession (where studies of this nature have been performed) seems appropriate.

A key difference found in studies comparing newly graduated medical students with experienced doctors is time. When time in making a decision is not a factor, "student recall will exceed experts" (Schmidt, Norman, & Boshuizen, 1989, p. 17). On the other hand, when time becomes a factor, "the trend reverse[s] and experts recalled more than novices (Schmidt, Norman, & Boshuizen, 1989, p. 17). An explanation of these phenomena is provided by Norman (1990) "since expert knowledge is compiled and [newly graduated students] are actively elaborating mechanisms, [newly graduated students] will recall more, but will require more time to process the text. Thus under conditions of unrestricted time. . .student recall will exceed experts." Other studies comparing clinical experience and expertise show no differences between experienced experts and newly graduated students (Feltovich, Johnson, Moller, & Swanson, 1984). Again, the key factor is time. Experienced professionals have the situations that they have seen in the past to act as templates for new situations that they see in the present. For example, a man comes to an experienced doctor with symptoms of vomiting and intestinal cramps. At the same time, the doctor notices that the man's skin has a yellow tinge. An experienced doctor might be able to relate the case to a similar set of symptoms from a person treated last month, last year, etc. An inexperienced doctor would need to start from scratch putting all of the symptoms together to diagnose the illness taking more time than the experienced doctor. Given time restrictions, an experienced professional performs better than an inexperienced professional in medicine.

The instrument in this study was used without time restrictions. If the medical explanation of experienced versus inexperienced differences is an applicable explanation for the field of instructional design, the time factor could be one explanation for the non-significant difference between mean scores of the professional masters and the non-professional masters.

Recommendations and Summary

At this point we would like to make recommendations regarding the use of this instrument and the future research of an instrument of this type.

A Research Tool

While the idea for the instrument came about from a continuing dialogue concerning certification in instructional design, it was and is not expected that this instrument be used in such a process. It is a research tool for use in a person's lifelong research agenda.

Further Research

Further research is needed for two reasons. First, many of the professional masters wrote helpful notes and suggestions for further item refinement as they completed items. It is felt from those responses that some changes in the instrument need to be made. Suggestions included grammatical changes to enhance question clarity and item content changes to reduce ambiguity. The second reason for further research is to broaden the scope of the subject groups. In the current study nonmaster subject groups were primarily students in education and science. Using the instrument with other professional groups would provide more information about the instrument's discriminating abilities. Other professional groups might include business administrators and professional trainers. Because of the overlapping competencies between those fields and instructional design, questions concerning how finely the instrument can discriminate could be addressed.

A paper and pencil, multiple-choice instrument can validly discriminate between masters and nonmasters of instructional design although further research is needed. The field of instructional design is quickly growing. As the field continues to grow and gain in importance, instructional designers will be expected to be more accountable for their abilities and actions. Questions of ability will become more prominent and questions such as the one in this study will need to be answered along with this higher demand for accountability.

Table 1
Professional Masters Demographic Data--Jobs Held

Job Held	Response Frequency
Corporate Educational Specialist	1
Corporate Executive	9
Corporate Instructional Developer	3
Faculty	16
Human Factors Specialist	1
Private Consultant	2

N = 32

Table 2
Professional Masters Demographic Data--Degrees Held

Degree	Response Frequency
Ph.D.	24
Ed.D.	5
M.A.	1
B.A.	1

N = 31

Table 3
Professional Masters Demographic Data--Program or Department in Which Degree Was Obtained

Program or Department	Response Frequency
Anthropology	1
Curriculum and Administration	1
Curriculum and Instruction	3
Education	2
Educational Psychology	1
Educational Technology	2
Instructional Design	2
Instructional Systems Technology	10
Instructional Technology	4
Psychology	4

N = 30

Table 4
Descriptive Statistics

Group	n	Maximum	Minimum	Mean	Standard Deviation
All Subjects	184	25	3	13.310	4.601
Professional Masters	34	25	14	19.265	2.632
Education Graduates	43	19	5	14.535	2.881
Non-Education Graduates	23	16	3	10.696	3.336
Education Undergraduates	45	19	9	13.622	2.516
Non-Education Undergraduates	39	13	4	7.949	2.523

Table 5
Tukey HSD Multiple Comparison Test for the Professional Master Group and the Four Nonmaster Subgroups

Group	Professional Masters	Education Grad.	Education Undergrad.	Non-Education Grad.	Non-Education Undergrad.
	N = 34	N = 43	N = 45	N = 23	N = 39
Mean	19.265	14.535	13.622	10.696	7.949

$p \leq .05$
Harmonic N = 34.720
Critical Difference = 1.813

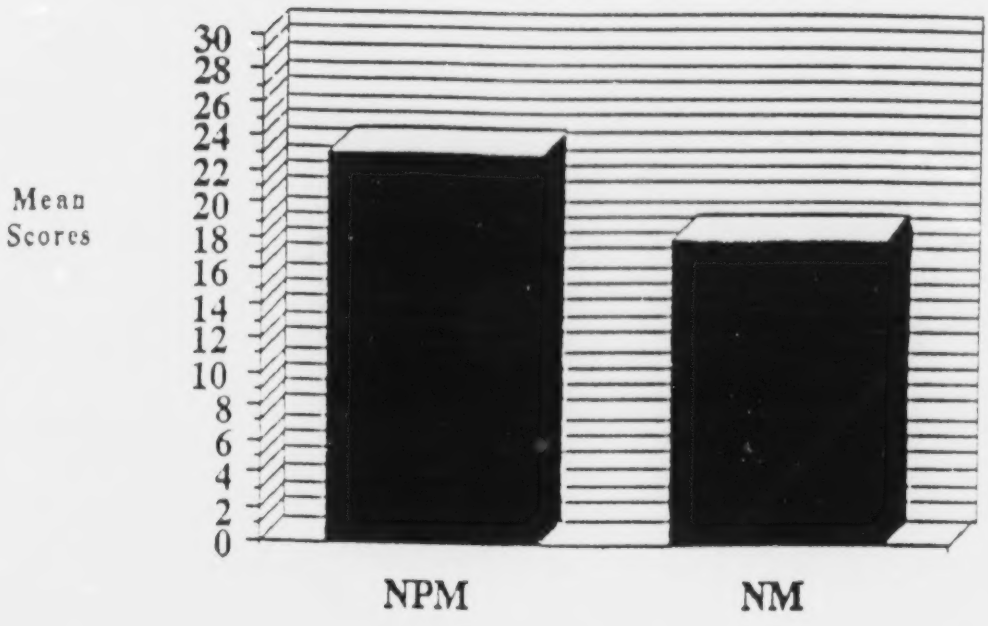
* Group or subgroup formed from professional master and nonmaster groups used in reliability and concurrent validity analysis.

Figure 1. Timeline showing three stages of the instrument's development.

Figure Caption

Figure Caption

Figure 2. Non-professional master and nonmaster mean scores in item analysis.

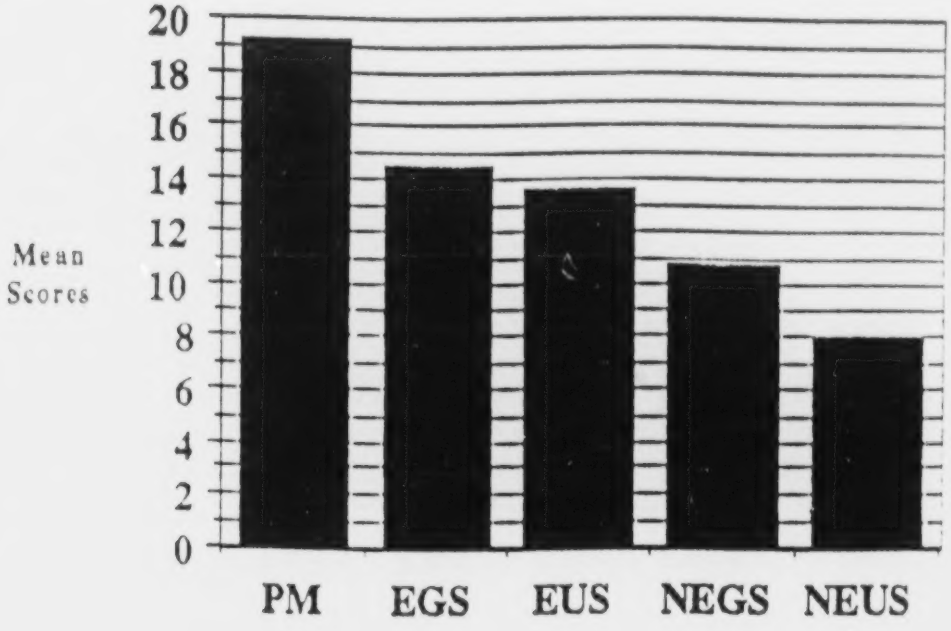


*NPM = Non-Professional Masters

NM = Nonmasters

Figure Caption

Figure 3. Professional master group and nonmaster subgroup mean scores.



*PM = Professional Masters

EGS = Education Graduates

EUS = Education Undergraduates

NEGS = Non-Education Graduates

NEUS = Non-Education Undergraduates

Figure Caption

Figure 4. Smoothed frequency distributions of masters and nonmasters.

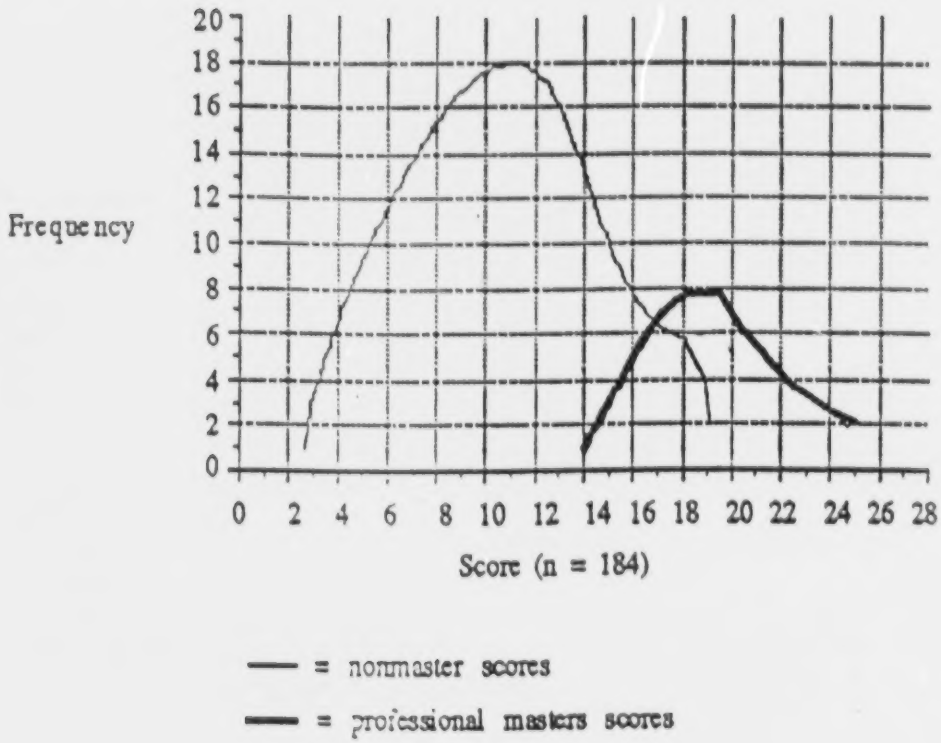


Figure Caption

Figure 5. Phi matrix at a cut off score of 17.

Classification of Instrument

Classification of Subject Matter Experts	Nonmaster	Master	
Master	4 11.76%	30 88.24%	4 + 30 = 34
Nonmaster	134 89.33%	16 10.67%	134 + 16 = 150
	4 + 134 = 138	30 + 16 = 46	Σ = 184

n = 184

Figure Caption

Figure 6. Actual and predicted frequencies produced by a discriminant analysis.

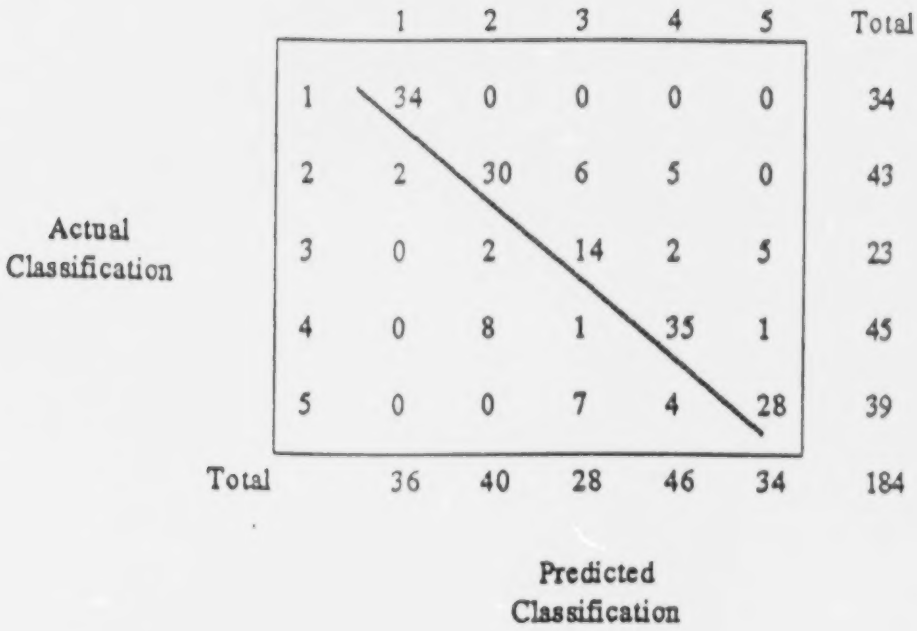
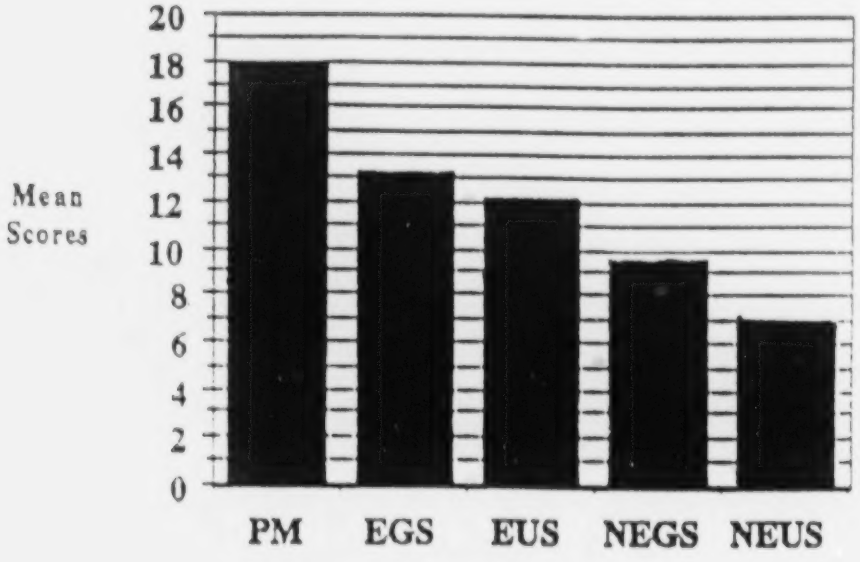


Figure Caption

Figure 7. Professional master and nonmaster subgroup mean scores (24 items).



*PM = Professional Masters

EGS = Education Graduates

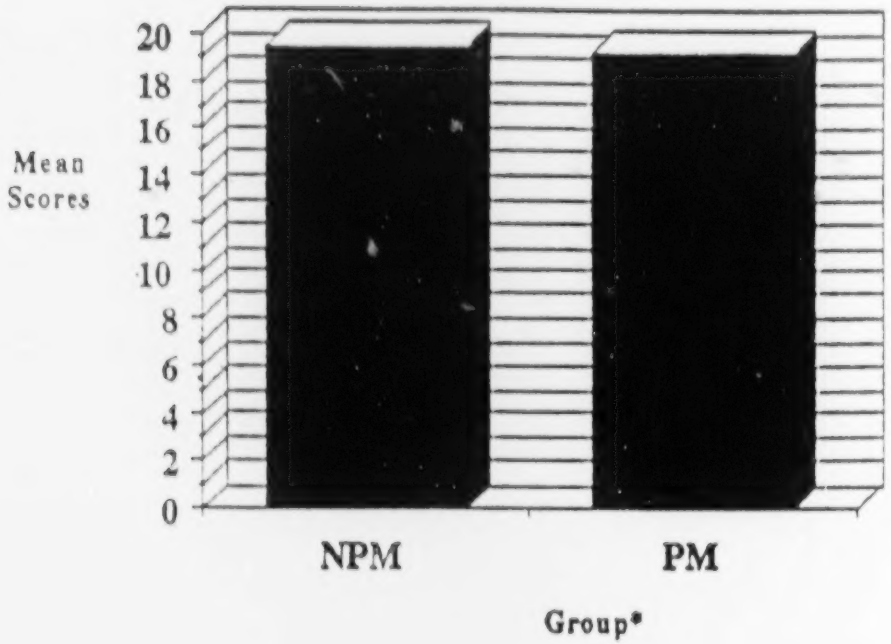
EUS = Education Undergraduates

NEGS = Non-Education Graduates

NEUS = Non-Education Undergraduates

Figure Caption

Figure 8. Non-professional master and professional master mean scores (28 items).



*NPM = Non-Professional Masters

PM = Professional Masters

Allen, M.J., & Yen, W.M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing Company.

Boothe, B. (1984). Certification--beyond reason. *Performance & Instruction Journal*, *23*(1), 19-20.

Bratton, B. (1984). Professional certification: Will it become a reality? *Performance & Instruction Journal*, *23*(1), 4-7.

Briggs, L.J. (Ed.). (1976). *Instructional design: Principles and applications*. Englewood Cliffs, NJ: Educational Technology Publications, Inc.

Bureau of Labor Statistics. (1986). *Occupational projections and training data: A statistical and research supplement to the 1986-87 occupational outlook handbook* (Bulletin 2251). Washington, DC: U.S. Government Printing Office.

Coscarelli, W. (1984). Arguments for certification: A political "trip-lett? *Performance & Instruction Journal*, *23*(1), 21-22.

Davis, R.H., Alexander, L.T., & Yelon, S.L. (1974). *Learning system design: An approach to the improvement of instruction*. New York: McGraw-Hill Book Company.

Deden-Parker, A. (1979). Instructional design competencies for business and industry designer-client interactions. *Educational Technology*, *19*(5), 44-46.

Deden-Parker, A. (1981). Instructional technology skills sought by industry. *Performance & Instruction Journal*, *20*(1), 24-25, 30.

Feltovich, P.J., Johnson, P.E., Moller, J.H., & Swanson, D.B. (1984). LCS: The role and development of medical knowledge in diagnostic expertise. In W.J. Clancy, & F. Shortlesse (Eds.). *Readings in Medical Artificial Intelligence: The past decade* (pp. 275-319). Reading, MA: Addison-Wesley.

Fleming, M., & Levie, W.H. (1978). *Instructional message design: Principles from the behavioral sciences* (4th ed.). Englewood Cliffs, NJ: Inc.

Gagne, R.M. (Ed.). (1987). *Instructional technology: Foundations*. Hillside, NJ: Lawrence Erlbaum Associates, Publishers.

Galey, M. (1980). Certification means professional development. *Instructional Innovator*, *25*(9), 25-31.

Glass, G.V., & Hopkins, K.D. (1984). *Statistical methods in education and psychology* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc.

International Board of Standards for Training, Performance, and Instruction. (1986). *Instructional design competencies: The standards*. Iowa City: International Board of Standards for Training, Performance, and Instruction.

Laird, D. (1985). *Approaches to training and development*. New York: Addison-Wesley Publishing Company, Inc.

Logan, R.S. (1982). *Instructional systems development: An international view of theory and practice*. New York: Academic Press.

Mager, R.F. (1973). *Measuring instructional intent or got a match?* Belmont, CA: Fearon Publishers.

Markle, S.M. (1978). *Designs for instructional designers*. Champaign, IL: Stipes Publishing Company.

McCullough, M. (1981). ASTD professional development update. *Training and Development Journal*, *35*(7), 17-18.

Moses, J., & Byham, W. (1977). *Applying the assessment center method*. New York: Pergamon Press.

Norman, G. (1990). *Research in the psychology of clinical reasoning: Implications for assessment*. Manuscript submitted for publication.

Pinto, P.R., & Walker, V.W. (1978). What do training and development professionals really do? *Training and Development Journal*, *32*(7), 58-64.

Prigge, W.C. (1974, November). Accreditation and certification: A frame of reference. *Audiovisual Instruction*, 12-18.

Romiszowski, A.J. (1981). *Designing instructional systems: Decision making in course planning and curriculum design*. London: Nichols Publishing.

Schmidt, H.G., Norman, G.R., & Boshuizen, H.P.A. *A cognitive perspective on medical expertise: Theory and implication*. Manuscript submitted for publication.

Shrock, S.A., & Foshay, W.R. (1984). Measurement issues in certification. *Performance & Instruction Journal*, *23*(1), 23-27.

Spaid, O.A. (1986). *The consummate trainer: A practitioner's perspective*. Englewood Cliffs: Prentice-Hall.

Stolovitch, H.D. (1981). Preparing the industrial and educational instructional developer: Is there a difference? *Performance & Instruction Journal*, *20*(1), 29-30.

Swezy, R.W. (1981). *Individual performance assessment: An approach to criterion-referenced test development*. Englewood Cliffs: Prentice-Hall.

Tabachnick, B.G., & Fidell, L.S. (1983). *Using multivariate statistics*. New York: Harper & Row.

Task Force on ID Certification. (1981). Competencies for the instructional/training development professional. *Journal of Instructional Development*, *5*(1), 14-17.

Tiemann, P.W., & Markle, S.M. (1978). *Analyzing instructional content: A guide to instruction and evaluation*. Champaign, IL: Stipes Publishing Company.

Trimby, M.J. (1982). *Entry-level competencies for instructional developers*. Dallas, TX: Paper presented at the Annual Meeting of the Association for Educational Communications and Technology. (ERIC Document Reproduction Service No. ED 222 174)

Wallington, D.J. (1981). Generic skills of an instructional developer. *Journal of Instructional Development*, *4*(3), 28-32.

Westgaard, O. (1983). Certification for performance and instruction professionals: The solution lies in the use of assessment centers. *Performance & Instruction Journal*, *22*(7), 3-7.

Westgaard, O. (1984). Certification in instructional design really is an issue. *Performance & Instruction Journal*, *23*(1), 3.

Appendix A
References Used to Identify Instructional Design Competencies

Analyzing instructional content: A guide to instruction and evaluation . . . . . Tiemann, P.W. & Markle, S.M.

Designing instructional systems: Decision making in course planning and curriculum design . . . . . Romiszowski, A.L.

Designs for instructional designers . . . . . Markle, S.M.

Individual performance assessment: An approach to criterion-referenced test development . . . . . Swezy, R.W.

Instructional design: Principles and applications . . . . . Briggs, L.J.(Ed.)

Instructional message design: Principles from the behavioral sciences . . . . . Fleming, M. & Levie, W.H.

Instructional systems development: An international view of theory and practice . . . . . Logan, R.S.

Instructional technology: Foundations . . . . . Gagne, R.M. (Ed.)

Learning system design: An approach to the improvement of instruction . . . . . Davis, R.H., Alexander, L.T., & Yelon, S.L.

Measuring instructional intent or got a match? . . . . . Mager, R.F.

The consummate trainer: A practitioner's perspective . . . . . Spaid, O.A.

Appendix B
Example Questions in the Instructional Design Assessment Instrument

1. All things being equal, which of the following concept definitions would trainees learn more easily?
   a. A desktop publishing program is a program for importing word processing files and for page layout and for mixing graphics and text.
   b. A microcomputer system error occurs from sloppy programming or there is a memory error or there is a hardware error.
   c. Refined oil is thicker than water but not quite as thick as crude oil.
   d. An incorrect formative evaluation is an evaluation that was not done or it was done incorrectly or it was only partially done.

2. You have to develop training for a group of secretaries who will have to use a new word processing program to be used by the company. The company has not used computers for word processing before this time and a survey has shown that most of the secretarial staff have never used a computer. What would be the most economical and efficient instructional sequence for each part of the instruction?
   a. statement of a step, example of the step, another example of the step requiring a secretary response
   b. example of a step, statement of a step, another example requiring a secretary response
   c. statement of a step, example of the step requiring a secretary response
   d. statement of a step, restatement of the step requiring a secretary response

3. A trainer is concerned that he is talking too quickly for the learners to understand and take good notes. What would be the most appropriate method of collecting data to see if the trainer's concerns are valid?
   a. an open-note test
   b. an audio recording of the trainer's lecture
   c. a final course evaluation completed by the learners
   d. observation of the training sessions by another trainer

4. The final evaluation for a required training unit to teach telephone operators to use a new long distance dialing system consists of a questionnaire. Below are 4 questions from the questionnaire. Which of the 4 questions should be eliminated?
   a. Was the training helpful?
   b. Did the instructor ask a lot of questions?
   c. Were the computer simulations useful in learning the task?
   d. Were the workbook exercises useful?